Principal component analysis and radiative transfer modelling of Spitzer Infrared Spectrograph spectra of ultraluminous infrared galaxies

Article (Published Version)

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **424,** 2069–2078 (2012) doi:10.1111/j.1365-2966.2012.21352.x

Principal component analysis and radiative transfer modelling of *Spitzer* Infrared Spectrograph spectra of ultraluminous infrared galaxies

P. D. Hurley,[1⋆] S. Oliver,[1] D. Farrah,[1,2] L. Wang[1] and A. Efstathiou[3]

[1]*Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH*
[2]*Virginia Polytechnic Institute & State University, Department of Physics, MC 0435, 910 Drillfield Drive, Blacksburg, VA 24061, USA*
[3]*School of Sciences, European University Cyprus, Diogenes St, Engomi, 1516 Nicosia, Cyprus*

Accepted 2012 May 21. Received 2012 May 11; in original form 2011 October 13

ABSTRACT

The mid-infrared spectra of ultraluminous infrared galaxies (ULIRGs) contain a variety of spectral features that can be used as diagnostics to characterize the spectra. However, such diagnostics are biased by our prior prejudices on the origin of the features. Moreover, by using only part of the spectrum they do not utilize the full information content of the spectra. Blind statistical techniques such as principal component analysis (PCA) consider the whole spectrum, find correlated features and separate them out into distinct components.

We further investigate the principal components (PCs) of ULIRGs derived in Wang et al. We quantitatively show that five PCs are optimal for describing the Infrared Spectrograph spectra. These five components (PC1–PC5) and the mean spectrum provide a template basis set that reproduces spectra of all $z < 0.35$ ULIRGs within the noise. For comparison, the spectra are also modelled with a combination of radiative transfer models of both starbursts and the dusty torus surrounding active galactic nuclei (AGN). The five PCs typically provide better fits than the models. We argue that the radiative transfer models require a colder dust component and have difficulty in modelling strong polycyclic aromatic hydrocarbon features.

Aided by the models we also interpret the physical processes that the PCs represent. The third PC is shown to indicate the nature of the dominant power source, while PC1 is related to the inclination of the AGN torus.

Finally, we use the five PCs to define a new classification scheme using 5D Gaussian mixture modelling and trained on widely used optical classifications. The five PCs, average spectra for the four classifications and the code to classify objects are made available at: http://www.phys.susx.ac.uk/pdh21/PCA/.

Key words: galaxies: statistics – infrared: galaxies.

1 INTRODUCTION

Ultraluminous infrared galaxies (ULIRGs) are galaxies whose rest-frame infrared luminosities, $L_{8-1000\,\mu m}$, exceed $10^{12}\,L_{\odot}$. Although ULIRGs were first discovered using ground-based photometry in the 1970s (Rieke & Low 1972), the *IRAS* survey transformed our understanding by observing the objects in much larger numbers (Soifer et al. 1984). Most have high star formation rates (SFRs $> 100\,M_{\odot}\,yr^{-1}$), while around half also contain an embedded active galactic nucleus (AGN).

ULIRGs are rare in the local Universe, with less than 50 at $z \lesssim 0.1$, but the associated luminosity function shows strong, positive evolution with redshift (e.g. Sanders 1999), resulting in several hundred ULIRGs per square degree at $z > 1$ (Rowan-Robinson et al. 1997; Barger et al. 1998; Hughes et al. 1998; Eales et al. 2000; Fox et al. 2002; Le Floc'h et al. 2005). The increase in number density with redshift and their associated high SFR means that ULIRGs make a significant contribution to the history of star formation at high z.

The mid- to far-infrared luminosity of ULIRGs is a result of dust and gas reprocessing the optical and ultraviolet (UV) radiation emitted by stars and/or AGN. Obtaining spectroscopy for the mid-infrared part of the spectrum became possible with instruments such as the *Infrared Space Observatory* (*ISO*; Kessler et al. 1996), and the Infrared Spectrograph (IRS; Houck et al. 2004) on the *Spitzer Space Telescope* (Werner et al. 2004). The ULIRG spectra from these instruments contain a wealth of spectral features. These include the emission lines from broad polycyclic aromatic hydrocarbons (PAHs), which are strong in star-forming regions, but absent

⋆E-mail: P.D.Hurley@sussex.ac.uk

in AGN-dominated sources (Moorwood 1986; Roche et al. 1991). A prominent [Ne v] 14.3 μm fine-structure line indicates the presence of an AGN, while the silicate features at 9.7 and 18 μm probe source geometry (Imanishi et al. 2007).

Combinations of the PAH emission lines, mid-infrared fine-structure lines and silicate features have been used as diagnostics for characterizing the power source behind the ULIRGs (Genzel et al. 1998; Rigopoulou et al. 1999; Farrah et al. 2007, 2008; Spoon et al. 2007; Farrah et al. 2009). There are however problems associated with these diagnostic tools, such as the separation of emission lines from both the continuum and underlying PAH features, the mixture of neighbouring features and different diagnostics giving conflicting estimates. They also only focus on small parts of the spectrum, disregarding the information contained in the remainder.

Larger regions of the spectrum can be investigated with the multivariate statistic, principal component analysis (PCA). PCA has been used for spectral classification for optical galaxies (e.g. Connolly et al. 1995; Bromley et al. 1998). Wang et al. (2011) carried out PCA on the IRS spectra of 119 local ULIRGs. They argued, qualitatively, that only four principal components (PCs) were needed to reproduce the variance in the ULIRG spectra. They also proposed that the contribution from each PC had some underlying physical interpretation. Examination of the first four PCs, and comparisons to the diagnostics employed by Spoon et al. (2007) and Nardini et al. (2009) suggested that PC1 constrains the dust temperature and geometry of the distribution of source and dust, while PC2 and PC3 determine the amount of star formation. The fourth PC is important for Seyfert type 2 galaxies, and is hence a possible indicator of an unobscured AGN.

In this paper we extend Wang et al. (2011) by quantitatively investigating how many PCs are needed to explain the variation in the spectra and compare the PC reconstructions to fits provided by a suite of radiative transfer models. We investigate what information the radiative transfer models are missing. We also re-examine what physical properties are behind the PCs, by investigating the relationship between the physical parameters of models and the contributions from different PCs. Finally, we introduce a new classification scheme using 5D Gaussian mixture modelling and trained with optical classifications. Section 2 gives an overview of the data and Section 3 gives a brief description of PCA. Section 4 will review the radiative transfer models being applied, and Section 5 will present the results. Conclusions will be presented in Section 6. We assume a spatially flat cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega = 1$ and $\Omega_\mathrm{m} = 0.3$.

2 THE DATA

This paper uses the same sample of mid-infrared spectra as Wang et al. (2011). We summarize their selection criteria here. The ULIRGs were observed as part of the IRS Guaranteed Time program (Armus et al. 2007; Farrah et al. 2007; Spoon et al. 2007) and those observed by Imanishi et al. (2007). An upper redshift cut of $z = 0.35$ was applied to ensure that we sample approximately the same wavelength range for each object. A further eight objects were removed as they have poor-quality data in the longer wavelength IRS module. In total, there are 119 objects in the sample.

3 PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA works by determining the eigenvectors from the covariance matrix of a given data set. For 119 spectra, each with 180 wavelength points, the 180 × 180 covariance matrix quantifies the correlation between each spectral point. The eigenvectors of the matrix can be thought of as spectral components that can be linearly combined to reconstruct each object in the sample.

Any spectrum can be linearly decomposed by projecting it on to the PCs defined by the 119 ULIRG sample. This allows each spectrum to be described by the contribution from each PC. These contributions define coordinates in a multidimensional space which we refer to as PCA space.

4 RADIATIVE TRANSFER MODELS

To compare with the fits provided by the PCs, we have carried out a minimum chi-squared (χ^2) search for linear combinations of a grid of starburst models described in Siebenmorgen & Krügel (2007) and grid of AGN dusty torus models of Efstathiou & Rowan-Robinson (1995). The libraries contain 5948 and 2109 spectral energy distributions (SEDs), respectively, and we have considered linear combinations of each AGN and starbust SED, giving us many models to search over.

4.1 Starburst models

We use the Siebenmorgen & Krügel (2007) starburst models. The models presented by Siebenmorgen & Krügel (2007) have been described as 'hotspot' starbursts. OB stars are assumed to be surrounded by dense clouds (the hotspots) and other stars, such as old bulge stars or massive stars, are dispersed in the diffuse medium. It is the hotspots that contribute to the mid-infrared part of the spectrum. The outer radius of these environments is determined by the condition of equal heating of the dust by the OB stars in the centre and the interstellar radiation field.

Both stellar groups are treated as continuously distributed sources, and the number density of both types of stars falls off as r^{-1}.

The parameters of these models include the starburst radius, R; ratio of the luminosity of OB stars with hotspots to total luminosity, f_OB; the total luminosity of the starburst, L_SB; total extinction from the outer radius of the galactic nucleus to its centre, A_v; and dust density of the hotspot environment, ρ_HS, corresponding to hydrogen number densities (n_HS) and assuming a gas-to-dust ratio of 150. The parameter ranges can be found in Table 1. In total, the library contains 5948 SEDs.

4.2 AGN torus models

This paper uses the AGN tapered disc models of Efstathiou & Rowan-Robinson (1995). The tapered disc models, in combination with the starburst models of Efstathiou, Rowan-Robinson & Siebenmorgen (2000), have been successful in fitting the SEDs of ULIRGs (Farrah et al. 2003), hyperluminous infrared galaxies (Farrah et al. 2002; Verma et al. 2002; Efstathiou 2006), submillimetre galaxies (Efstathiou & Siebenmorgen 2009) and active galaxies (Alexander

Table 1. Parameter values and ranges for the starburst models.

Parameter	Range
R (kpc)	0.35, 1 and 3
f_OB	0.4, 0.6 and 0.9
L_SB ($\mathrm{L_\odot}$)	10^{10}–10^{14} in steps of 0.1 dex
A_v (mag)	2.2, 4.5, 7, 9, 18, 35, 70 and 120
n_hs ($\mathrm{cm^{-3}}$)	10^2, 10^3, 2.5×10^3, 5×10^3, 7.5×10^3 and 10^4

Table 2. Parameter values and ranges for the AGN models.

Parameter	Range
τ	500, 750, 1000 and 1250
Θ (°)	30, 45 and 60
r_{in}/r_{out}	20, 60 and 100
θ (°)	0–90 with either
	40 or 75 divisions (depending on r_{in}/r_{out})

et al. 1999; Ruiz et al. 2001; Efstathiou & Siebenmorgen 2005; Farrah et al. 2012). The torus is modelled as a disc, whose thickness increases with distance from the central source but tapers off in the outer regions of the torus. The dust density is distributed smoothly within the disc and follows a r^{-1} relation, with r being radius. The parameters for the AGN torus model are as follows: UV equatorial optical depth to the centre of the torus, τ; the opening angle of the torus, Θ; the ratio of inner to outer radius of the torus, r_{in}/r_{out}; and the viewing angle, θ. The parameter ranges can be found in Table 2. In total, there are 2109 AGN SEDs.

4.3 The fitting procedure

We have considered all linear combinations of a starburst and AGN model when fitting the observed spectra of the 119 ULIRG sample. We use the wavelength grid of the starburst models, and the lower resolution AGN models are interpolated on to the same grid. The smoothness of the AGN models makes the interpolation justifiable. The radiative transfer models lack molecular hydrogen emission so we mask out regions of the spectrum where molecular hydrogen features occur (i.e. 9.46–9.86, 12.08–12.48 and 16.83–17.23 μm).

The wavelength resolution of the PCs is higher than the starburst model resolution. For proper comparison to the fits, and to allow decomposition of the models into PCA space, we have re-derived the PCs for the ULIRG sample at the resolution of the starburst models. There is no significant change in the shape of components. We also note that the sign of the PC contributions for each object remains the same and the change in magnitude of the PC contributions is not significant in comparison to the spread of contributions for the sample.

To remain consistent with the analysis of Wang et al. (2011), the models are normalized so that the mean flux over the whole wavelength range is unity.

We then carry out a linear least-squares fit for each combination of starburst and AGN model, with the condition that the fit parameters are positive (i.e. to eliminate the possibility of a negative amount of starburst or AGN). Model comparison is then carried out via minimum χ^2.

We assumed a minimum of 5 per cent flux error for each spectral bin of the IRS spectra, which is consistent with the observed variations between individual nod positions on the IRS as described in Chapter 7 of the IRS Instrument Handbook.[1]

5 RESULTS

5.1 Optimum number of components

We first investigate how many PCs are needed to describe the ULIRG sample. Wang et al. (2011) did not quantitatively show whether four PCs were sufficient. Using the PCs rederived at the lower resolution described in Section 4.3, we have investigated how many PCs are needed to accurately reconstruct the IRS spectra of all 119 ULIRGs in the sample. For each spectrum, we quantify the goodness of reconstruction with the reduced chi-squared statistic χ^2_ν, where the number of degrees of freedom is equal to the number of wavelength points minus the number of PCs used in the reconstruction.

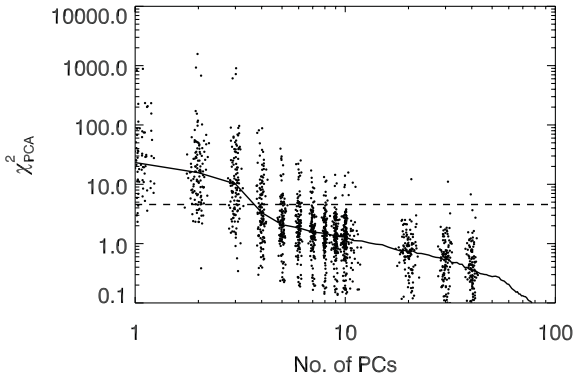

Figure 1. The median variation of the χ^2_ν for the PC reconstruction as the number of components used in the reconstruction is increased. The dashed line indicates the median χ^2_ν for the radiative transfer model fits. For PC reconstructions using up to 10 PCs (and 20, 30, 40), we also plot the χ^2_ν for every object (offset for clarity).

Fig. 1 shows that as we increase the number of PCs used in the reconstruction, the median χ^2_ν value for the sample decreases. We have plotted the χ^2_ν for each individual object for reconstructions using up to 10 PCs and the median χ^2_ν value obtained by fitting the ULIRGs with the radiative transfer models as described in Section 4.3. 10 PCs would appear to be the optimal number, i.e. where $\chi^2_\nu = 1$. We find that four PCs (assumed by Wang et al. 2011) give a median χ^2_ν of 3.3, while adding a fifth component substantially decreases the median χ^2_ν to 2.1. The use of six and seven PCs only reduces the median χ^2_ν to 1.8 and 1.6, respectively.

The eigenvalues associated with each PC are a measure of the variance each PC accounts for and provide an alternative method to determine the optimum number of components. In Fig. 2, we

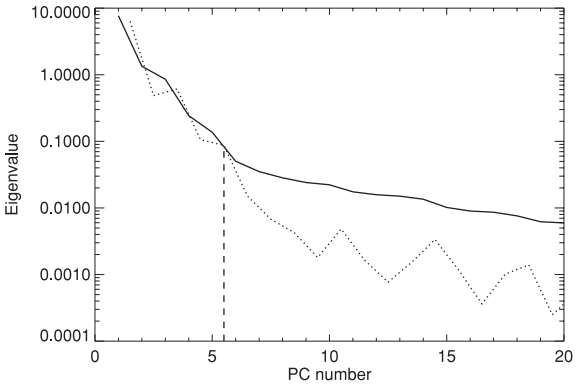

Figure 2. The eigenvalues (solid line) and difference in eigenvalues (dotted line) for the PCs. The eigenvalues quantify the variance associated with each PC, and are a measure of importance. The difference between eigenvalues drops dramatically for the first few PCs, but levels off beyond five (indicated by the dashed line). We therefore argue that five PCs are a more suitable number than the four used in Wang et al. (2011).

[1] http://irsa.ipac.caltech.edu/data/SPITZER/docs/irs/

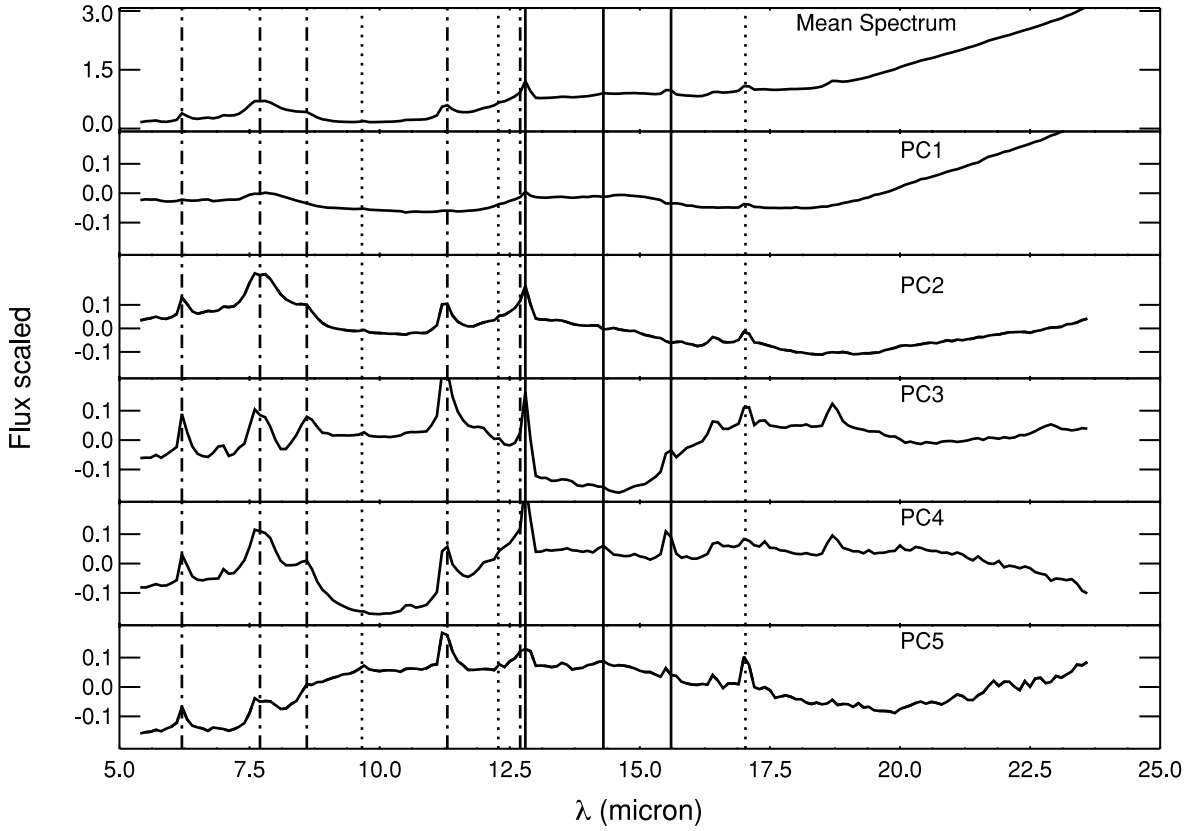

Figure 3. The mean spectrum and principal components for the sample of ULIRGs. The dot–dashed vertical lines mark the central location of the 6.2, 7.7, 8.6, 11.2 and 12.7 μm PAH emission lines. The dotted lines indicate the location of the molecular hydrogen lines at 9.66, 12.28 and 17.03 μm. The solid vertical lines indicate the position of the neon fine-structure lines: [Ne II] 12.8, [Ne V] 14.3 and [Ne III] 15.6 μm.

plot the eigenvalues and difference in eigenvalues for the PCs. The general trend indicates that the difference between eigenvalues significantly decreases with each component. The exception to the rule occurs between the third–fourth component and the fifth–sixth component where the difference between eigenvalues is larger than the trend. We associate this larger than expected difference as in indication that the previous component captures significantly more information than the next. This suggests that the third and fifth PCs are substantially more important than the fourth and sixth, respectively. Beyond the sixth PC, the trend flattens out, indicating that most of the variation related to structure has been captured. Overall, Figs 1 and 2 do not definitively indicate the optimum number of PCs. However, we argue that the reduction in χ^2_ν to 2.1 and difference in eigenvalue between the fifth and sixth PCs indicate that five PCs rather than the four PCs used by Wang et al. (2011) strike a better balance of providing a small basis set of templates, whilst adequately describing the spectra.

The fifth component was not discussed in Wang et al. (2011) and so we now show this component, compared to the original four. The mean spectra of the 119 ULIRGs and the five components can be seen in Fig. 3. There are a number of spectral features in this fifth component, most notably the 6.2, 11.2 and 12.7 μm PAH emission lines as well as the molecular hydrogen emission line at 17.03 μm. The 6.2 μm emission feature has negative flux, while the 11.2 and 12.7 PAH lines are both positive. Overall, the fifth component does not contain any new features that were not seen in the previous components. Its role appears to be in altering the ratios of existing features.

5.2 Analysis of the radiative transfer models

We now investigate whether the radiative transfer models discussed in Section 4 are capable of modelling the spectra. An example of the fit produced by five PCs and the radiative transfer models can be seen in Fig. 4.

We now compare all the χ^2_ν for reconstructions using five PCs with the χ^2_ν for our radiative transfer model fits. Fig. 5 shows the

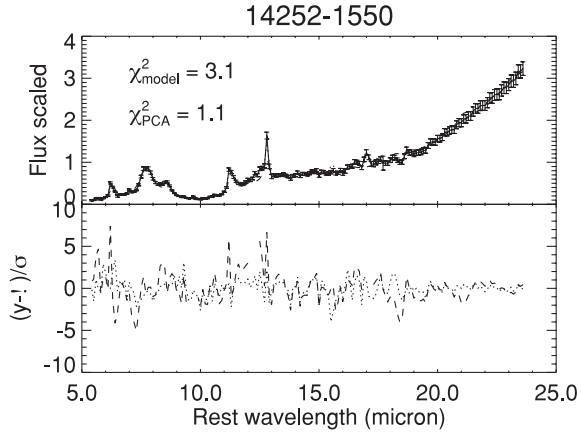

Figure 4. An example of our fit with 14252−1550. The radiative transfer model is plotted with a dashed line, and the principal component reconstruction with five PCs is shown with a dotted line. The residual over error is also shown to indicate where either technique may be failing

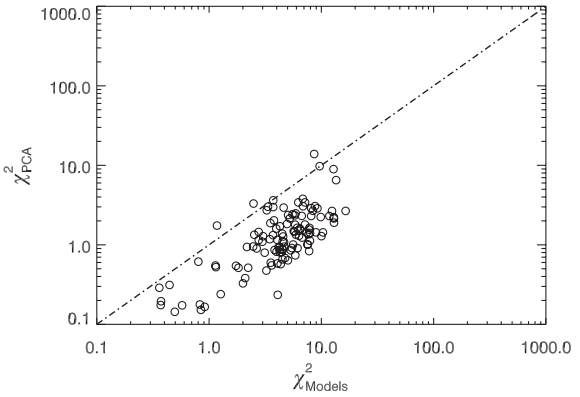

Figure 5. The χ^2_ν values for each object in the sample for both radiative transfer model fits and the five-PC reconstruction. Most objects do better with the PCs.

distribution of the reduced χ^2 values for both the five-PC reconstructions and the radiative transfer model fits, for all ULIRGs in the sample. A five-component reconstruction fits the spectra better, on average, than the radiative transfer models.

We have shown that five PCs can explain the sample of ULIRGs better than the radiative transfer models, but the two are not competing methodologies. The PCs will always do better than the models as they are derived from the data and the number of PCs is increased until the reproduction of the spectra is good. They represent an extraction of most of the important information from the spectra. Radiative transfer models are used to give us physical information of objects. However, Fig. 5 indicates that the ULIRGs are not modelled well on average by the radiative transfer models.

By comparing the models to the PCs, we can investigate what information is in the PCs that are not in the models. Fig. 6 shows the contributions made by the five PCs, as a function of χ^2_ν for the model fits. We only plot objects that have a reasonable χ^2_ν for the five-PC reconstruction, i.e. a $\chi^2_\nu \leq 3$. We also bin the model χ^2_ν values into three bins. The mean and 1σ dispersion are overplotted as filled circles and error bars.

The general increase of $\chi^2_{\nu\rm Models}$ in Fig. 6 shows that models tend to do worse when the objects have a large, positive contribution from PC1. Wang et al. (2011) suggested a large, positive contribution from PC1 indicated colder dust. Our results suggest that objects with colder dust are not well modelled by the AGN and starburst component models. The increase in dispersion with $\chi^2_{\nu\rm Models}$ for PC2 and PC3 indicates that models do worse when there is a large, absolute contribution from PC2 and PC3. PC2 and PC3 relate to strong spectral lines, which would indicate that the models have problems with constraining the strength of spectral lines. Lower values of $\chi^2_{\nu\rm Models}$ appear to occur when objects have negative values of PC4, but as the $\chi^2_{\nu\rm Models}$ values increase beyond 2, there appears to be little change in PC4. A negative contribution in PC4 would suppress emission features, indicating that models are again inadequate in modelling spectral features. There appears to be little change of PC5 contribution with $\chi^2_{\nu\rm Models}$.

In Fig. 7, we show the stacked difference between spectra and radiative transfer model fits (solid line) and the spectra and five-PC reconstructions (dotted line). The stacked difference for spectra and models illustrates that model fits underestimate the PAH spectral lines and do not include Neon fine-structure lines, or molecular hydrogen lines. The PAH underestimate is consistent with our interpretation of Fig. 6. It suggests that the Kruegel (2003) PAH treatment used by the Siebenmorgen & Krügel (2007) starburst models is unsuitable for the extreme star-forming ULIRGs. As expected, the PC reconstructions perform considerably better than the models.

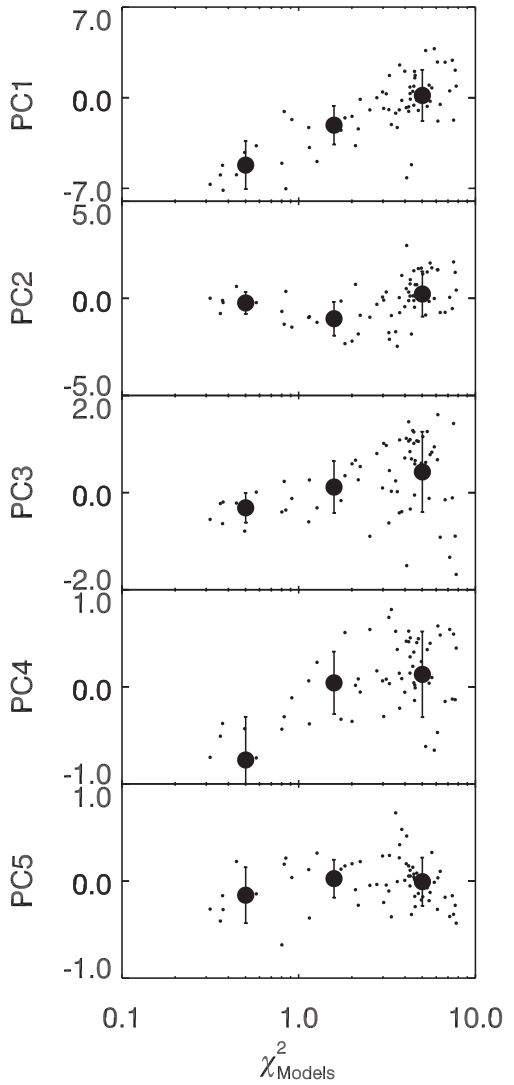

Figure 6. The contributions made by each PC against the χ^2_ν from the model fits. Only objects with a five-PC fit of $\chi^2_\nu \leq 3$ have been plotted. The mean and 1σ dispersion for three bins are overplotted as filled circles and error bars.

5.3 Interpreting the principal components

We have shown that five PCs provide a simple empirical basis set that captures most of the important variations in ULIRGs. We have also shown some limitations of the models. Nevertheless, the models still describe some of the physics of the objects and can be cautiously used to investigate whether the components are associated with physical parameters. We investigate the components by directly comparing the PC contributions and the radiative transfer model best fits for the ULIRG sample. Fig. 8 shows the contribution from each PC as a function of the viewing angle and starburst/AGN contribution. We have binned the PC contributions and calculated the average and 1σ dispersion for each bin. These are overplotted with error bars. PC1 shows a correlation with viewing angle of AGN, with positive contributions corresponding to an obscured AGN and negative to face-on AGN. The contribution from the fourth

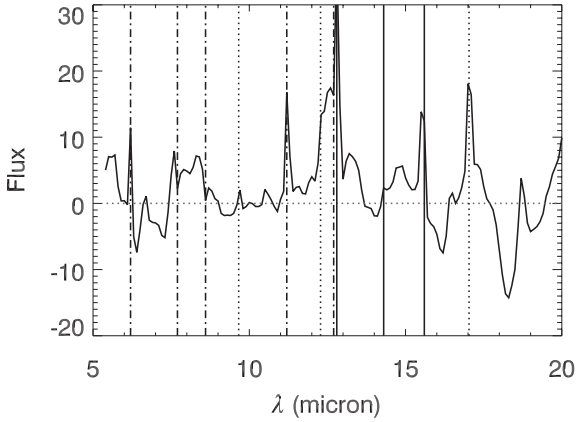

Figure 7. The stacked difference between ULIRG spectra and best-fitting radiative transfer models (solid line) and the ULIRG spectra and five-PC reconstructions (dotted line). The dot–dashed vertical lines mark the central location of the 6.2, 7.7, 8.6, 11.2 and 12.7 μm PAH emission lines. The dotted lines indicate the location of the molecular hydrogen lines at 9.66, 12.28 and 17.03 μm. The solid vertical lines indicate the position of the neon fine-structure lines: [Ne II] 12.8, [Ne V] 14.3 and [Ne III] 15.6 μm.

PC appears to drop with viewing angle from around $\pi/4$ radians. The other PCs show no discernible dependence. The starburst/AGN contribution is plotted against PCs in the right-hand side of Fig. 8. Negative values of PC3 seem to be associated with AGN-dominated sources and positive values with starbursts. The other PCs show a large amount of dispersion and little correlation with starburst/AGN contribution.

We now decompose the radiative transfer model fits into the PCA space described in Section 3. The position of each ULIRG (squares) in four of the PCA planes and corresponding best-fitting model (filled circles) can be seen in Fig. 9. The mean difference between the ULIRGs and models is depicted by the arrow in the top right of each plane.

We find that the location of radiative transfer model best fits in PCA space is offset relative to the ULIRG positions. There are numerous explanations for the offset. The sparseness of the model library could be a factor. The decomposition into PCA space may also be affected by the missing physics in the models. We therefore treat the model tracks with caution and limit interpretation to relative changes in PC contribution rather than absolute position.

We have taken the best-fitting radiative transfer model and vary each parameter in turn to see how it affects the position in PCA space. We focus on the viewing angle of AGN and ratio of starburst to AGN power, which we define as

$$L_{\rm total} = pL_{\rm SB} + (1-p)L_{\rm AGN}. \quad (1)$$

A value $p = 0$ describes a pure AGN model, and $p = 1$ relates to a complete starburst.

Fig. 10 shows the 1D parameter tracks for 50 randomly selected ULIRGs. The viewing angle tracks show a decrease in PC1 contribution when going from an obscured to face-on AGN. Tracks in PC4 are curved, indicating a non-linear relationship with viewing angle. PC3 appears to be a good indicator for the power ratio, with PC3 contribution decreasing as AGN power begins to dominate. Tracks in PC5 also show a slight correlation with power ratio, while for the other PCs the relationship is unclear.

The interpretation of tracks is consistent with the conclusions drawn from Fig. 8. Certain PCs appear to be related to the physics of the ULIRGs. We have shown that PC1 is linked to AGN viewing angle, while PC3 is linked to the star formation and AGN contribution. This is consistent with the interpretation of Wang et al. (2011).

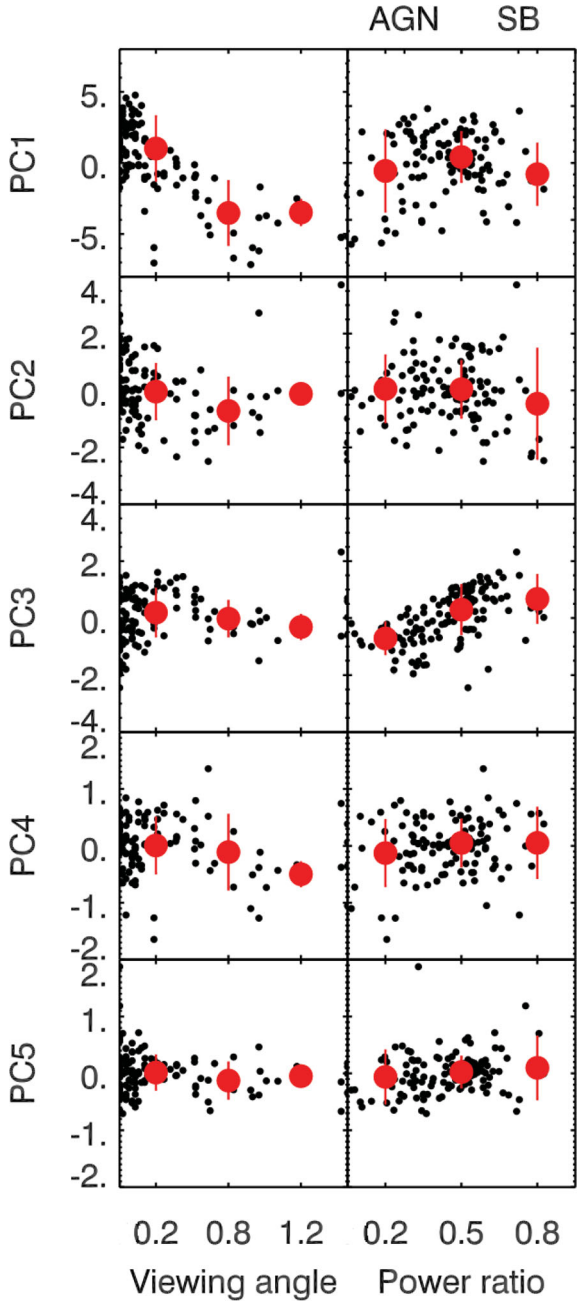

Figure 8. The contribution from each PC against the radiative transfer parameters of viewing angle and starburst/AGN contribution. The average contribution for three bins and associated 1σ dispersion are overplotted.

5.4 Gaussian mixture classification scheme

Since we have shown that the PCs capture most of the information in IRS spectra, it is natural to use the PCs as a classification tool. Wang et al. (2011) suggested that position in the PC1–PC4 plane was related to optical type. We now take this one step further by proposing a classification scheme based on optical classifications, using the multidimensional Gaussian mixture modelling applied in Davoodi et al. (2006). This type of parametric modelling works by assuming that the density function of galaxies in our 5D PCA space

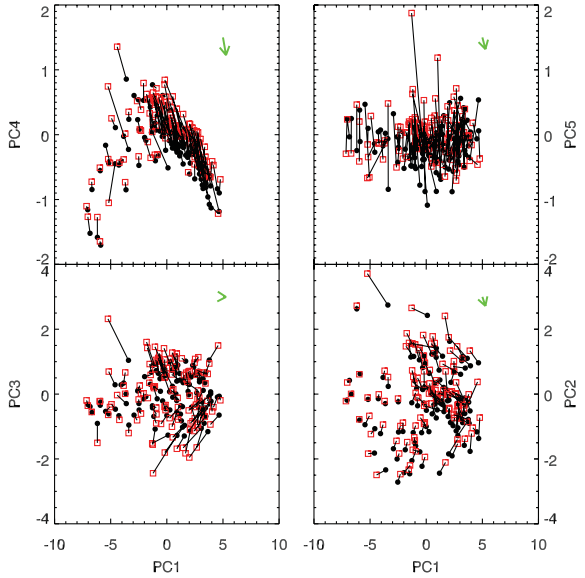

Figure 9. The position of ULIRGs in four of the PC planes (squares) and the position in PCA space of the corresponding best-fitting radiative transfer models (filled circles). Each ULIRG and best-fitting model are joined by a solid line. The arrows in the top left of each plot show the mean difference between ULIRGs and models.

is composed of a mixture of multidimensional Gaussian functions. We take the four optical classifications (Seyfert 1, Seyfert 2, low-ionization nuclear emission-line region galaxy (LINER) and H II) that exist for 78 of our 119 ULIRG sample, and assume that the density of objects in each classification can be described as Gaussian. The resulting position and width of each Gaussian are trained from the optical classifications. They can be thought of as a probability density function (PDF) that describes the probability of belonging to each optical classification, as a function of position in PCA space.

Fig. 11 shows the marginalized 1σ contours for the optical classifications in four 2D projections. We note that the 1σ contours are for visualization only, our classification scheme makes use of all five dimensions. The objects with optical classifications are represented with different symbols: crosses for Seyfert 1, triangles for Seyfert 2, squares for LINER and open circles for objects classified as H II. Objects without an optical classification are plotted with a diamond. The success rate of our classification can be found in Table 3.

The classification scheme is very successful in correctly identifying Seyfert 1-like objects, while most of the Seyfert 2s are classified correctly of as LINERs. The majority of LINER objects are correctly identified, while the majority of H II optically classified ULIRGs are spread across H II and LINER groups. Both the LINER and the H II classifications lie in similar areas of PCA space, and discrete classification for objects in this region may not be completely appropriate as many ULIRGs will show signs of both. Overall our 5D Gaussian classification scheme works well in associating regions in PCA space with type of object and is a powerful tool in objectively classifying objects.

We have used our classification scheme to classify the 41 ULIRGs with no optical classification. The percentages can be seen in Table 3. We find that the majority are H II and LINER objects, while 12 per cent are classified as Seyfert 2-like objects. None of the objects appears to be Seyfert 1, suggesting that optical classification of Seyfert 1 objects is complete. We now make use of our 5D Gaussian classification scheme by creating average spectra for our four classifications using all 119 ULIRGs. Before averaging

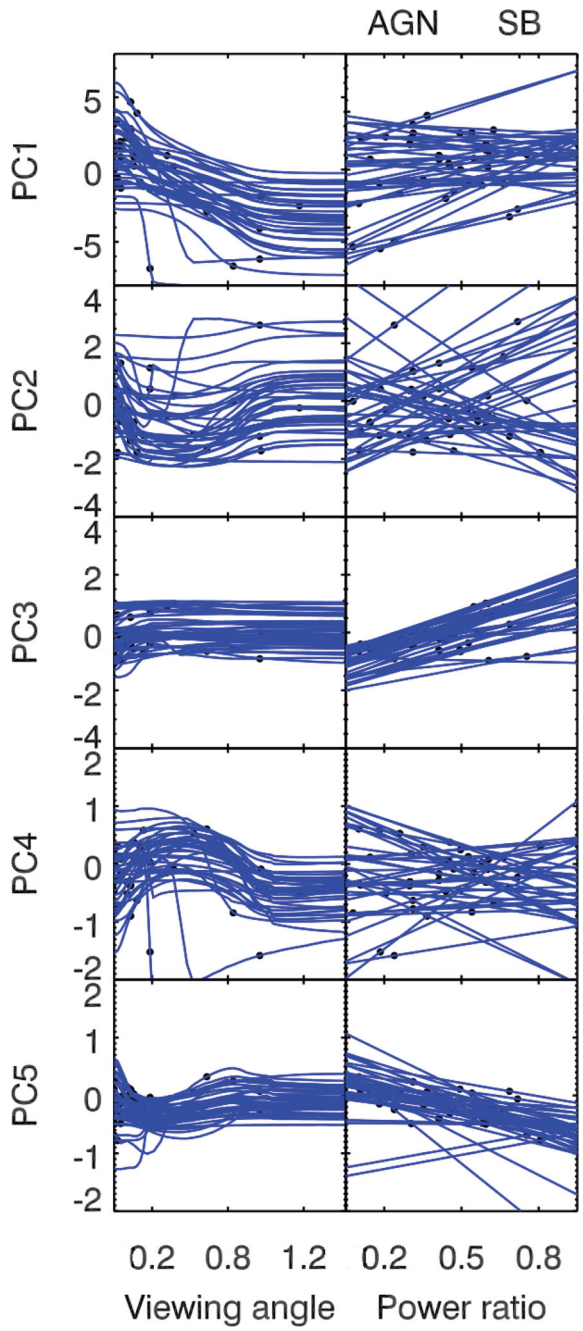

Figure 10. The contribution from each PC against viewing angle tracks and power ratio for 50 of the ULIRGs. For each best-fitting radiative transfer model, the viewing angle and power ratio have been varied to create tracks in PCA space.

the spectra, each spectrum is normalized so that the mean flux over the whole wavelength range is unity. The resulting four average templates can be seen in Fig. 12. As expected, the H II and LINER templates are similar, whilst Seyfert templates have very little PAH emission.

6 CONCLUSIONS

We have shown that five PCs are needed to describe most of the variation in the 119 local ULIRG sample and are more successful than a full χ^2 fitting by radiative transfer models. We have examined

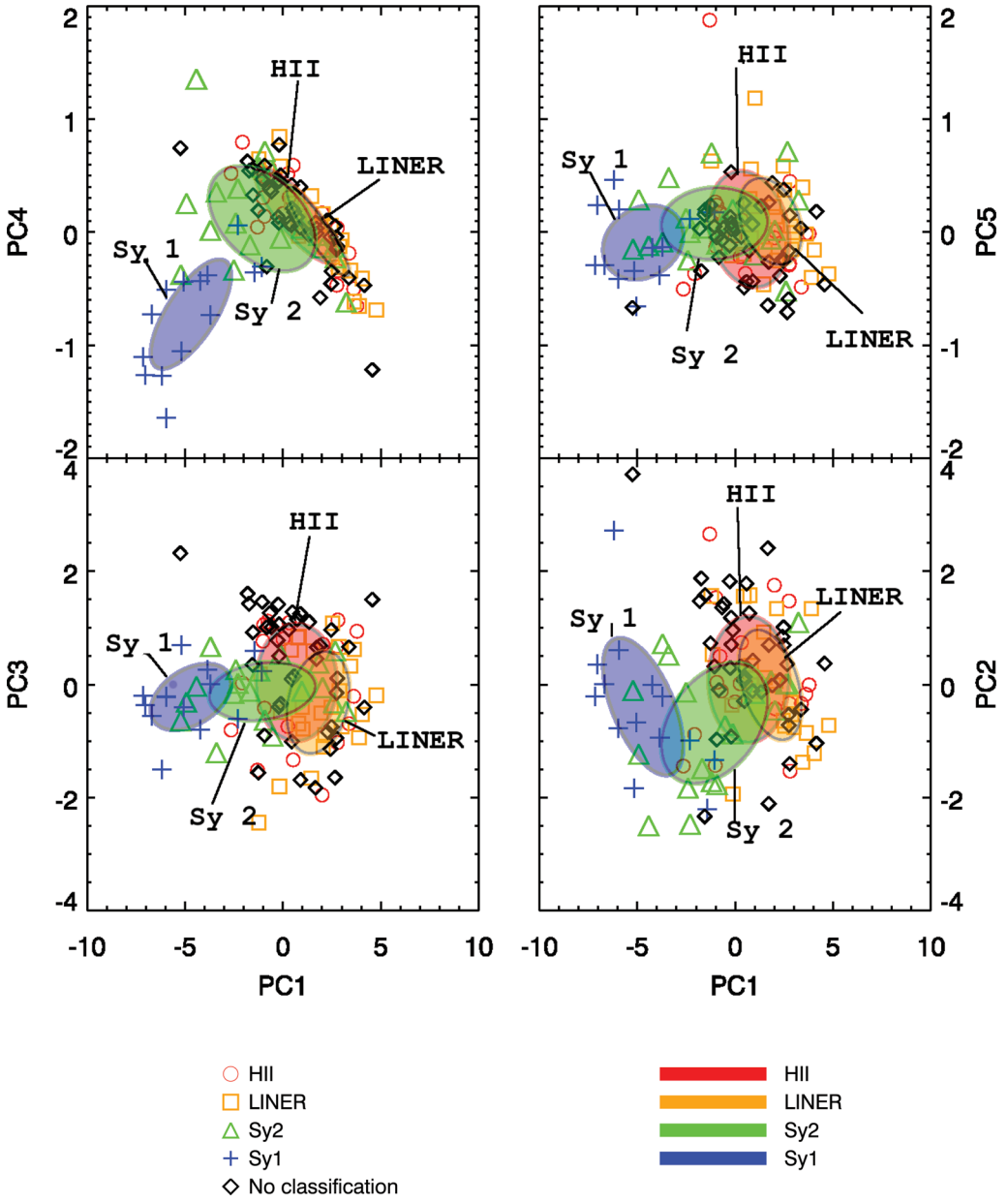

Figure 11. Four out of the possible 10 2D projections for our PCA space with the 1σ contours for the Gaussian mixture based classifications. Optically classified Seyfert 1 objects are marked by crosses, Seyfert 2 by triangles, LINERs by squares and H II classified objects with open circles. Those objects without optical classification are marked by diamonds.

Table 3. The percentage of objects in the four classifications as a function of their original classification. Not classified refers to those objects without an optical classification.

		Gaussian classification			
		H II (per cent)	LINER (per cent)	Sy 2 (per cent)	Sy 1 (per cent)
Optical	H II	40	45	15	0
	LINER	11	85	4	0
	Sy 2	5	26	51	18
	Sy 1	0	0	6	94
Not classified		42	46	12	0

what the radiative transfer models are missing. The fits provided by radiative transfer models appear to need a cold dust component and have difficulty in modelling the strength of strong PAH emission lines.

We have used a combination of the Siebenmorgen & Krügel (2007) starburst models and Efstathiou & Rowan-Robinson (1995) AGN torus templates to investigate what physical parameters are behind the components. We have examined how best-fitting model parameters are related to PC contribution. Overall, our conclusions are consistent with those reached in Wang et al. (2011). Contributions from PC1 appear to indicate the viewing angle of AGN with negative contributions associated with face-on AGN and positive for obscured AGN. PC3 appears to be the best indicator of whether it is the AGN or starburst that is the prevailing power source.

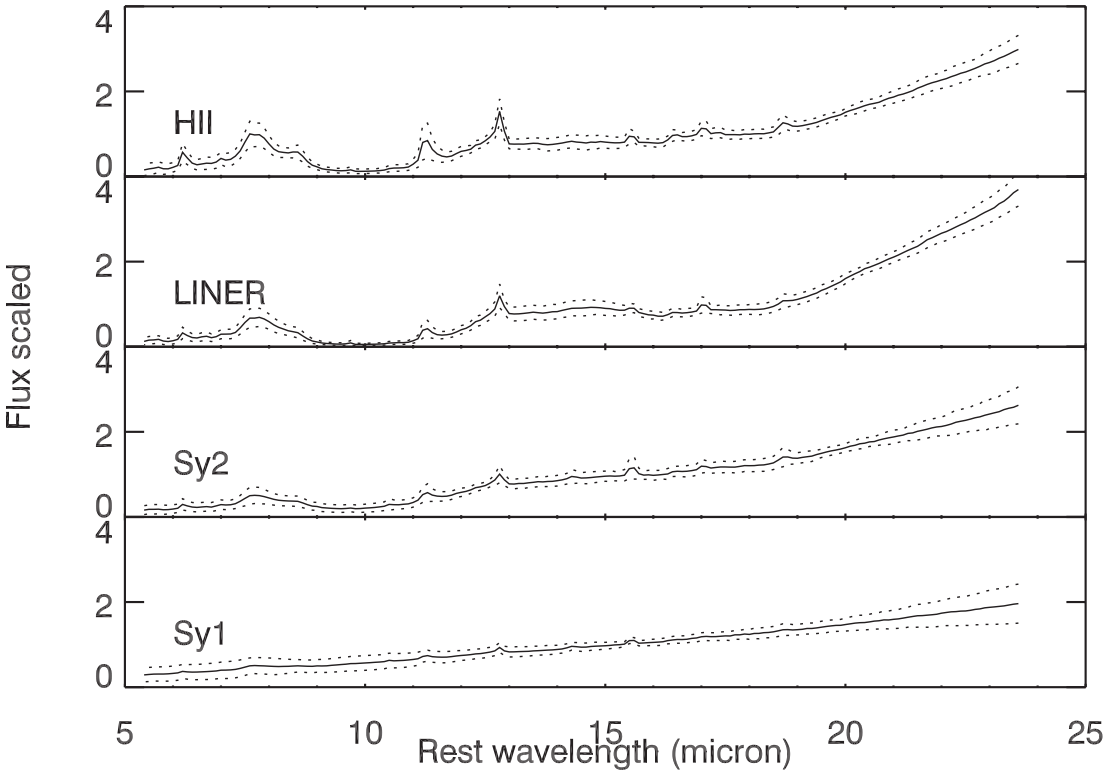

Figure 12. The average spectra for the four classifications, H II (29 objects), LINER (54 objects), Seyfert 2 (20 objects) and Seyfert 1 (16 objects). The dotted lines represent the 1σ dispersion in each classification.

The PCs consider a large part of the mid-infrared spectrum and are therefore less likely to be affected by problems associated with diagnostics based on single spectral features such as the PAH emission lines, where measuring line strength can be difficult. We suggest that the five PCs would be useful as empirical templates for ULIRG spectra in the IRS public data base (Lebouteiller et al. 2011).

We also introduce a new Gaussian mixture classification scheme based on location in the 5D PCA space and trained via optical classifications. Objects can be classified as either Seyfert 1, Seyfert 2, LINER or H II-like. We note that any ULIRG with IRS spectra (in the relevant wavelength range) can be decomposed on to the PCs, and the position in PCA space can be used to classify the object.

We have used our classification scheme to provide a set of average spectra for the four groups. We make these, the five PCs and code to classify objects available at: http://www.phys.susx.ac.uk/pdh21/PCA/.

ACKNOWLEDGMENTS

We acknowledge support from the Science and Technology Facilities Council (grant numbers ST/F006977/1, ST/I000976/1 and PP/E005306/1). This work is based on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. We thank the referee for the very helpful comments which led to improvements in this paper.

REFERENCES

Alexander D. M., Efstathiou A., Hough J. H., Aitken D. K., Lutz D., Roche P. F., Sturm E., 1999, MNRAS, 310, 78
Armus L. et al., 2007, ApJ, 656, 148
Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H., 1998, Nat, 394, 248
Bromley B. C., Press W. H., Lin H., Kirshner R. P., 1998, ApJ, 505, 25
Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D., 1995, AJ, 110, 1071
Davoodi P. et al., 2006, AJ, 132, 1818
Eales S., Lilly S., Webb T., Dunne L., Gear W., Clements D., Yun M., 2000, AJ, 120, 2244
Efstathiou A., 2006, MNRAS, 371, L70
Efstathiou A., Rowan-Robinson M., 1995, MNRAS, 273, 649
Efstathiou A., Siebenmorgen R., 2005, A&A, 439, 85
Efstathiou A., Siebenmorgen R., 2009, A&A, 502, 541
Efstathiou A., Rowan-Robinson M., Siebenmorgen R., 2000, MNRAS, 313, 734
Farrah D., Serjeant S., Efstathiou A., Rowan-Robinson M., Verma A., 2002, MNRAS, 335, 1163
Farrah D., Afonso J., Efstathiou A., Rowan-Robinson M., Fox M., Clements D., 2003, MNRAS, 343, 585
Farrah D. et al., 2007, ApJ, 667, 149
Farrah D. et al., 2008, ApJ, 677, 957
Farrah D. et al., 2009, ApJ, 700, 395
Farrah D. et al., 2012, ApJ, 745, 178
Fox M. J. et al., 2002, MNRAS, 331, 839
Genzel R. et al., 1998, ApJ, 498, 579
Houck J. R. et al., 2004, ApJS, 154, 18
Hughes D. H. et al., 1998, Nat, 394, 241
Imanishi M., Dudley C. C., Maiolino R., Maloney P. R., Nakagawa T., Risaliti G., 2007, ApJS, 171, 72
Kessler M. F. et al., 1996, A&A, 315, L27
Kruegel E., 2003, The Physics of Interstellar Dust. IoP Publishing, Bristol (Institute of Physics)
Le Floc'h E. et al., 2005, ApJ, 632, 169
Lebouteiller V., Barry D. J., Spoon H. W. W., Bernard-Salas J., Sloan G. C., Houck J. R., Weedman D. W., 2011, ApJS, 196, 8
Moorwood A. F. M., 1986, A&A, 166, 4

Nardini E., Risaliti G., Salvati M., Sani E., Watabe Y., Marconi A., Maiolino R., 2009, MNRAS, 399, 1373
Rieke G. H., Low F. J., 1972, ApJ, 176, L95
Rigopoulou D., Spoon H. W. W., Genzel R., Lutz D., Moorwood A. F. M., Tran Q. D., 1999, AJ, 118, 2625
Roche P. F., Aitken D. K., Smith C. H., Ward M. J., 1991, MNRAS, 248, 606
Rowan-Robinson M. et al., 1997, MNRAS, 289, 490
Ruiz M., Efstathiou A., Alexander D. M., Hough J., 2001, MNRAS, 325, 995
Sanders D. B., 1999, Ap&SS, 269, 381
Siebenmorgen R., Krügel E., 2007, A&A, 461, 445
Soifer B. T. et al., 1984, ApJ, 278, L71
Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V., 2007, ApJ, 654, L49
Verma A., Rowan-Robinson M., McMahon R., Efstathiou A., 2002, MNRAS, 335, 574
Wang L., Farrah D., Connolly B., Connolly N., LeBouteiller V., Oliver S., Spoon H., 2011, MNRAS, 411, 1809
Werner M. W. et al., 2004, ApJS, 154, 1

This paper has been typeset from a TEX/LATEX file prepared by the author.